Thomas J. Watson Research Center, Yorktown Heights

SHORTEST PATH, ASSIGNMENT AND TRANSPORTATION PROBLEMS

by

A. J. Hoffman*
H. M. Markowitz**

3/14/63

1. Introduction

The purpose of this note is to add some insight into the already known relationships among the three problems mentioned in the title. In considering the shortest path through a network from some initial vertex to a terminal vertex, we shall confine ourselves to those cases in which the sum of the lengths of the edges around any cycle is nonnegative. Most, though not all, algorithms for solving the shortest path problem make such a presumption.

That the shortest path problem may be posed in the format of an assignment problem is well known, at least in folklore, and we will, for the sake of completeness, indicate how this is done. Further, that transportation problems may be solved by performing a succession of shortest path problems is also well known, the general principle being that expounded on Page 121 of the monograph [1]. What we shall

* The work of this author was supported in part by the Office of Naval Research under Contract No. Nonr 3775(00), NR 047040.

** The Rand Corporation

IBM

show is that, for the case of assignment and transportation problems, one can be even more stringent in specifying the succession of shortest path problems to be solved than what the principle expounded in [1] permits. For example, we shall show that one can do an $n \times n$ assignment problem by solving a succession of shortest path problems on vertices specified in advance.

2. The Assignment Problem

We assume that the assignment problem is given in the following form:

We are required to minimize

$$(2.1) \qquad \sum_{i,j} c_{ij} x_{ij}$$

where

$$X = (x_{ij})$$

is a square matrix of order n, with nonnegative entries, and

$$\sum_j x_{ij} = \sum_i x_{ij} = 1.$$

Now, let us assume that we have a directed graph with vertices $1, \ldots, n+1$, where the distance from i to j is a real number d_{ij}, these numbers satisfy the cycle condition mentioned in the introduction, and we wish to find the shortest path from vertex 1 to vertex $n+1$. We set up an assignment problem where the rows correspond to the vertices from 1 to n, the columns correspond to the vertices from 2 to n + 1, and the "c_{ij}" are as indicated in the following diagram:

$$(2.2)\qquad \begin{array}{c|c|c|c|c|c|}
 & 2 & 3 & \cdots & n & n+1 \\ \hline
1 & & & & & \\ \hline
2 & 0 & & d_{ij} & & \\ \hline
\cdot & & 0 & & & \\ \hline
\cdot & & & & & \\ \hline
\cdot & & & & & \\ \hline
n & & & & 0 & \\ \hline
\end{array}$$

The reason is as follows: It should be clear that the given assignment problem of order n is essentially the same as the assignment problem of order n + 1 given in the following diagram:

$$\begin{array}{c|c|c|c|c|c|}
 & 1 & 2 & \cdots & n & n+1 \\ \hline
1 & \infty & & & & \\ \hline
2 & \infty & 0 & d_{ij} & & \\ \hline
\vdots & \cdot & & 0 & & \\ \hline
 & \cdot & & \ddots & & \\ \hline
n & \infty & & & 0 & \\ \hline
n+1 & 0 & \infty & \cdots & \infty & \infty \\ \hline
\end{array}$$

Now, in order to solve the larger assignment problem, we seek a permutation matrix of order n + 1 whose inner product with the given matrix is as small as possible. In the larger problem, the minimizing permutation can be expressed as a product of disjoint cycles. Further, it is apparent from the large matrix that the cycle which contains 1 must return to 1 from n + 1, thereby picking out the sum of distances along some path from 1 to n + 1. Furthermore, all other cycles may be taken to be degenerate because the condition on the sum of the d_{ij} in any cycle tells us that it is optimal to make all other cycles consist of each of one element, thereby incurring an additional "c_{ij}" of 0. Thus, we see that solving the larger assignment problem is, on the one hand, equivalent to finding the shortest path and, on the other hand, equivalent to solving the smaller assignment problem.

Now, we propose to show how this process can be reversed, in a sense. Suppose we begin with an arbitrary assignment problem (2.1), and let us assume that we have solved the assignment problem of order r $(1 \leqq r < n)$ corresponding to the lower left $r \times r$ sub-matrix. Assume, for the sake of ease of notation, that the minimizing permutation occurred on the main diagonal of that sub-matrix. Then, on the $n \times n$ matrix C, perform the following operations:

Subtract $c_{n-r+1,1}$ from the $(n-r+1)^{th}$ row of C,

subtract $c_{n-r+2,2}$ from the $(n-r+2)^{th}$ row of C,

$\dots,$

subtract $c_{n,r}$ from the n^{th} row of C.

As is well known, this operation does not change the original assignment problem. Furthermore, if we consider the assignment problem of order $r + 1$ given by the lower left-hand square of order $r + 1$ in the new matrix, it has the appearance of (2.2), and appears to be a shortest path problem on $r + 1$ vertices. Furthermore, the cycle condition is satisfied; otherwise, we would not have solved the assignment problem of order r.

In this way, one can solve an assignment problem of order n, by successively solving shortest path problems of smaller order.

3. Transportation Problems

We now consider the application of a similar idea to transportation problems. We shall assume (with no loss of generality) that the transportation problem is given in the following form: Minimize $\sum_{i,j} c_{ij} x_{ij}$, where $C = (c_{ij})$ is an $m \times m$ matrix of given constants, $a_1, \dots, a_m$ are given nonnegative integers, $n = \Sigma a_1$, and $X = (x_{ij})$ satisfies $x_{ij} \geqq 0$, $\sum_j x_{ij} = a_{ij}$, $\sum_i x_{ij} = 1$.

The intuitive idea behind this method is to treat each column of X successively, and dispose of the unit x_{ij} to be disposed of in that column in the most economical way. Now, we can certainly begin in this fashion until, for some i, as many as a_i of the columns have been "assigned" to row i. After that, it may turn out that the cheapest assignment of some subsequent column may also be in row i, and it will not then be possible to complete the intuitive scheme. What will replace the intuitive scheme is the solution of a shortest path problem, involving at most $m + 1$ points. Then, we shall show that one can use the solution to this problem to modify the matrix of C in such a way that it will appear that the selected elements are still minimal in their respective columns.

To explain this in adequate detail, let us assume that the first k columns have been disposed of, and let $\bar{c}_{ij}$ $(j = 1, \ldots, k)$ be the least (and selected) c_{ij} in its respective column. For ease of notation, let us also assume that a_1 elements have been selected in row $1, \ldots,$ a_t elements have been selected in row t, but fewer than a_i elements have been selected in row i, $i = t + 1, \ldots, m$. We now define a shortest path problem on $t + 2$ vertices. In order to define the problem, we must give the distances between the points. First,

$$d_{0i} = c_{i,k+1} \quad \text{for} \quad i = 1, \ldots, t,$$

$$d_{0,t+1} = \min_{i > t} c_{i,k+1}.$$

Further, for $t + 1 > i > 0$, $t + 1 > j > 0$, we define $d_{ij} = \infty$ if row i has never been selected and $d_{ij} = \min\limits_{k \geq p \geq 1} c_{jp} - \bar{c}_{ip}$. It is understood that the only candidates in this minimization occur for those differences corresponding to columns p where row i has been selected. Note that, since all distances not involving the point 0 are positive, the cycle condition is satisfied. Finally, for

$$t + 1 > i > 0,$$

$$d_{i,t+1} = \min_{j>t} \min_{k \geq p \geq 1} c_{jp} - \bar{c}_{ip},$$

if row i has ever been selected, otherwise infinity.

Now, determine the shortest path from 0 to the point $t + 1$. For most methods of determining the shortest path (see, for example, pp. 130 ff. of [1]), one derives as well the shortest distance from 0 to any point. Let $\Pi_0 = 0, \Pi_1, \ldots, \Pi_{t+1}$ be the shortest distance from 0 to each point. Subtract Π_1 from the first row of $C, \ldots,$ subtract Π_t from the t^{th} row of C. Subtract Π_{t+1} from the remaining rows of C.

Also, make the following adjustments in selected elements: If the shortest path from 0 to $t + 1$ is the path $i_0 = 0, i_1, i_2, \ldots, i_r = t + 1$, and if

$$d_{i_1, i_0} = c_{i_1, k+1},$$

$$d_{i_1, i_2} = c_{i_2 p_1} - \bar{c}_{i_1 p_1},$$

$$d_{i_2, i_3} = c_{i_3 p_2} - \bar{c}_{i_2 p_2},$$

$$d_{i_r, i_{r-1}} = c_{j p_{r-1}} - c_{i_{r-1} p_{r-1}},$$

then select $c_{i_1, k+1}$ to represent the $k + 1^{th}$ column, and change the elements selected in column $p_1, p_2, \ldots, p_{r-1}$ from $c_{i_1 p_1}, c_{i, p_2}, \ldots, c_{i_{r-1} p_{r-1}}$ to $c_{i_2 p_1}, c_{i_3 p_2}, \ldots, c_{j p_{r-1}}$, respectively.

To prove the validity of this change, we must show that, after the matrix C has been transformed in the manner described, and after some of the selected elements have been changed in the manner described, then the new selected elements are still minimal in their columns, and that, for each row i, the i^{th} row has been chosen no more than a_i times. The last stipulation is, of course, obvious, so let us prove the first. To do this, it is sufficient to show that every previously selected element is still minimal in its column, and that every newly selected element is also minimal in its column. To show that every previously selected element is minimal, let us use the fact that the distances Π_i must satisfy the inequality (*) $\Pi_i + d_{ij} \geq \Pi_j$. Suppose $\bar{c}_{ip}$ were selected and c_{jp} any other element in the p^{th} column. We have (**) $d_{ij} \leqq c_{jp} - \bar{c}_{ip}$. In the transformed matrix, the new elements in position (i, p) and (j, p) respectively are $c_{ip} - \Pi_i$ and $c_{jp} - \Pi_j$. We must show

$$(***)\ c_{ip} - \Pi_i \leqq c_{jp} - \Pi_j,$$

but

$$c_{ip} - \Pi_i \leqq c_{ip} - \Pi_j + d_{ij} \quad \text{by } (*)$$

$$\leqq c_{ip} - \Pi_j + c_{jp} - c_{ip} \quad (\text{by } (**)\ )$$

$$= c_{jp} - \Pi_j.$$

To show that the new selected elements are also minimal in their respective columns, let us first consider the $(k + 1)^{th}$ column. It is clear that the selected element will be 0 after the transformation, and all other elements will be nonnegative, by virtue of (*) with $i = 0$. For the other columns where the selected elements have been changed, note that (*) and (**) will be equations, when c_{jp} is a newly selected element, so that (***) will also be an equation.

This completes the discussion.

Reference

[1] Ford, L. R., Jr. and D. R. Fulkerson, "Flows in Networks", Princeton University Press, 1962.

LIMITED DISTRIBUTION NOTICE

This report has been submitted for publication elsewhere and has been issued as a Research Note for early dissemination of its contents. As a courtesy to the intended publisher, it should not be widely distributed until after the date of outside publication.

Printed in U.S.A.